HOW TO MAKE SENSE OUT OF RESEARCH DATA

by
H. Harold Toba
Yakima Agricultural Research Laboratory
USDA, ARS, Yakima, Wa. 98902

Purpose of Experiment

The typical problem that confronts an agricultural researcher is how will the use of a new or different practice affect the outcome of some particular segment of agricultural enterprise, and to what extent. To answer such a problem, an experiment is generally required to discover something unknown or to test a principle or hypothesis. The experiment may simply involve the comparison of two treatments (the new practice and the old one), a more complicated one which includes several rates or methods of applying the new practice, or a complex one in which the effects of several practices are studied simultaneously. The purpose of conducting an experiment is to provide a means of making observations (collecting data) resulting from the application of the treatment. These observations are used to make generalizations about the practice under study. Results from various experiments are then used to make deductive conclusions, such as making predictions, deciding on preventive measures, and applying control measures.

Planning the Experiment

The most important phase of research is proper planning. The first step is to find out what is already known. This is followed by what needs to be known, i.e., establishing the objectives of the experiment. The success of an experiment in meeting the objectives depends on the careful selection of treatments, the collection of data, and the experimental design to answer the questions posed.

In designing an experiment, there are certain "rules" to follow, particularly if the observed data are to be analyzed statistically. Essential to all experimental design are the following three R's:

1. Replication, which means that a treatment is repeated two or more times. Its function is to provide an estimate of experimental error and to provide a more precise measure of treatment effects. The data cannot be analyzed statistically without replication.

2. Randomization, which means the assignment of treatments to experimental units (e.g. plots) so that all units have an equal chance of receiving a treatment. Its function is to assure unbiased estimate of treatment effects and experimental error.

3. Reference, which means something that the treatment can be compared against. A control, or check, is added to an experiment to account for effects by factors other than the treatment. Another reference is a standard, such as an old practice, which is used to compare against a treatment, such as a new practice.

In planning an experiment, much thought should go into deciding what will be observed or measured because this is what will be used to evaluate the treatment. If certain crucial observations are not obtained, then the experiment has to be repeated. This might be the reason why you sometimes hear the old saying, "the experiment asked more questions than it answered."

This Presentation is part of the Proceedings of the 1983 Washington Potato Conference and Trade Fair.

Finally, how will the data be analyzed? This, as mentioned previously, depends to a large extent on how the experiment was designed. Because there are several ways in which to statistically analyze the data, and each has a certain set of "rules" to follow, the experiment must be designed accordingly.

Execution of Experiment

Execution of an experiment is merely the application of what has been planned, which involves the laying out of the experiment as designed, the application of the treatments, and the collection of data. However, when laying out the experiment, care must be taken so as not to introduce bias, such as the exertion of certain factors selectively on experimental units. For example, an edge effect or the influence of adjacent plots on each other can be rectified by the inclusion of border rows.

No matter how well an experiment is planned, it may be necessary to make modifications due to unforeseen circumstances. For example, if you were applying a spray on a plot and you were unsure whether the plot received the amount it was supposed to, you are justified to discard that replicate from the experiment.

The data, which is the payoff of the experiment, must be accurately collected. When collecting data the experimental design is used so that differences among individuals or differences associated with order of collection can be included in the experimental error. Also, all observations should be recorded, which can later be used when interpreting the results.

Treatment of Data

When the experiment is completed, the observed data are summarized. Sometimes, differences between treatment means are obvious. However, if they are not, then they are analyzed statistically, which takes into account the observed data and the experimental design.

Statistics is a tool used by the researcher to analyze and interpret numerical data collected. It deals with observed data which were obtained under a certain set of conditions, and makes it possible to express the results in simple and logical terms. It does not, however, prove that the data or experiment is right or wrong.

Probability is an important aspect of statistical analysis. The term "at $p=0.05$" or "at the 5% level of probability" means that there is a 5% probability that the observed variations among means could occur by chance. When the observed variations among means are real, we say that the means are significantly different, keeping in mind that there is a 5% probability that they could have occurred by chance. No difference among the means does not prove that some of the treatments had no effect. There is always a definite probability that there was a real effect but that the experiment was too insensitive to detect the difference at the 5% level of probability.

The treatments and treatment means, statistically analyzed, comprise the results of the experiment, which are usually arranged in tables, or represented graphically.

Interpretations and Conclusions

The final step in the investigation consists of interpreting the data which have been collected. What are the conclusions growing out of the analyses? What do the figures tell us that is new or different, or that reinforces or casts doubt upon previous hypotheses? The results must be interpreted in the light of the limitations of the original data. Remember that the data themselves are but approximation; therefore, too exact conclusions must not be drawn from them. However, all of the useful applicable meaning which is present in the data should be discovered and clarified.

One of the greatest dangers in drawing conclusions is putting too much weight on numerical differences instead of statistical differences. Take for example an experiment in which different treatments were tested to determine their effect on the yield (tons/acre) of a crop. Example 1 shows that the mean yield for each of the treatments was significantly different from each other at the 5% level of probability. This is expected because all of the variability is treatment effects and no variability due to replicate effects. Example 2, however, shows no significant differences between the same means because of the tremendous variability due to replicates effects. Example 3 shows some variability due to replicates effects yet with significant separation for same treatment means, but not as much as in Example 1. Finally, Example 4 shows some variability in the replicates within treatments, but the variability between treatments are proportional to each other. Thus, more separation of the same means than in Example 3, but less than in Example 1.

Effect of various treatments on the yield (ton/acre) of a crop.

Treatment	I	II	III	IV	Total	Mean*
EXAMPLE 1						
A	1	1	1	1	4	1 a
B	2	2	2	2	8	2 b
C	3	3	3	3	12	3 c
D	4	4	4	4	16	4 d
EXAMPLE 2						
A	0	0	2	2	4	1 a
B	4	4	0	0	8	2 a
C	6	0	6	0	12	3 a
D	0	8	0	8	16	4 a
EXAMPLE 3						
A	0	0	2	2	4	1 a
B	0	0	4	4	8	2 ab
C	0	0	6	6	12	3 ab
D	0	0	8	8	16	4 b
EXAMPLE 4						
A	.3	.7	1.3	1.7	4	1 a
B	.6	1.4	2.6	3.4	8	2 ab
C	.9	2.1	3.9	5.1	12	3 bc
D	1.2	2.8	5.1	6.8	16	4 c

* Means followed by the same letter within a test are not significantly different at P=0.05, Duncan's multiple range test.

When looking at these four examples, one might ask, "which is right?" Remember, as mentioned previously, there is no right or wrong. Each of these results was the outcome of an experiment conducted under its respective, unique set of circumstances. This brings up another danger, and that is making inferences based on one experiment, especially under field conditions. Unless the various factors that could affect the outcome of a treatment are known to be similar, it would be risky to assume that the treatment will produce the same effect under all conditions. For example, you cannot expect to control an insect with a certain treatment when the population differs from that of the experiment, or to expect a certain yield when the soil type differs from that of the experiment. Also, as mentioned previously, there is still a 5% probability that the differences among the observed means, whether significant or not, could be due to chance. Therefore, under field conditions, it would be safer to either define the conditions under which a treatment is expected to perform, or to make conclusions based on the results of several experiments conducted under different conditions.

Another way to assess the effects of treatments is by correlation, which means the tendency of two variables to be related in a definite manner. Take the question, "How is the amount of applied fertilizer related to the yield of crop?" The two variables are the amount of fertilizer applied and the yield of the crop. They are called independent (X) and dependent (Y) variables, according to which one is viewed as depending on the other. When an increase in one variable is accompanied by an increase in the other, it is called direct or positive correlation (Fig. 1). On the other hand, when an increase in one variable is accompanied by a decrease in the other, it is called inverse or negative correlation (Fig. 2).

The closeness of the relationship can be calculated and expressed as the coefficient of correlation ($\underline{r}$). Since 1 is perfect correlation and 0 is no correlation, the higher the $\underline{r}$-value the higher the correlation. A test of significance measures what the probability is that such a correlation could be due to chance. The fewer the number of observations the higher the $\underline{r}$-value needs to be in order to be significant. An * following the $\underline{r}$-value means that the observations obtained would occur less than 5% of the time by chance alone, and ** for less than 1% of the time (Fig. 3). Also, observations must be made on at least three units of the X-variable, because if only two observations are made, there will always be a straight line and a perfect correlation.

The nature of the relationship is expressed by regression, which means the amount of change in one variable associated with a unit change in the other variable. When an increase in one variable is accompanied by a proportional increase or decrease in the other, it is called a linear regression (Figs. 1A and 2A). Conversely, when an increase in one variable is accompanied by an unproportional increase or decrease in the other, it is called a curvilinear regression (Figs. 1B and 2B). The degree of relationship in a linear regression is expressed by $\hat{Y} = \underline{a} + \underline{b}X$, where $\hat{Y}$ is an estimate of the dependent variable, $\underline{a}$ is the intercept or the point on the Y-axis at X=O, $\underline{b}$ is the slope or the steepness of the line, and X is the independent variable. Thus, it could be said that starting with Y units (the value of $\underline{a}$), every unit increase in X is associated by an average increase (or decrease) of $\underline{b}$ units of Y (Fig. 3).

There are certain pitfalls to be aware of when interpreting the data. (1) A low correlation does not always mean lack of relation. The data could fit a curvilinear relation. (2) A high correlation doesn't necessarily mean a cause and effect relationship. The correlation merely shows how the two variables are related to each other. (3) In a linear regression, extrapolation of a line beyond the range of observations should be avoided (Fig. 3). It is dangerous because the observations could have been made in only a short section of a line in which the variables were related in a curvilinear fashion (Fig. 4).

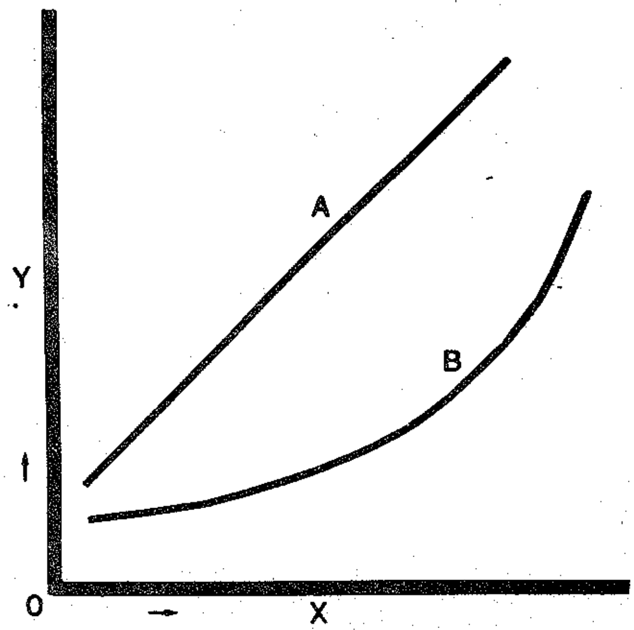

Fig. 1. Positive correlation

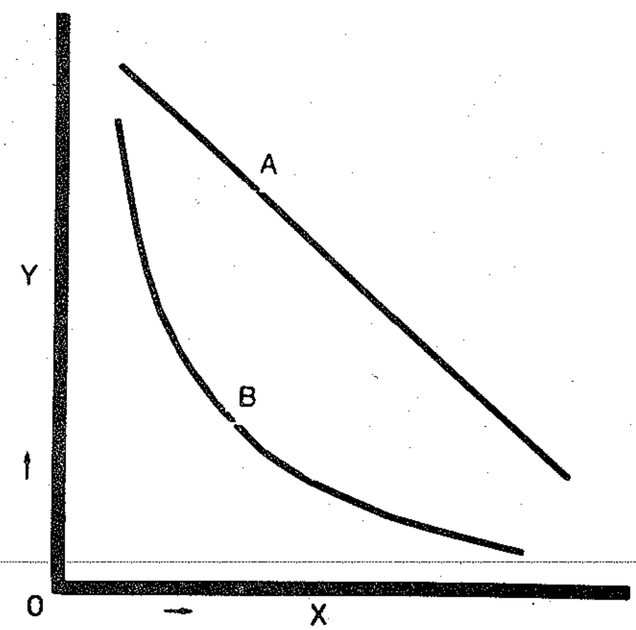

Fig. 2. Negative correlation

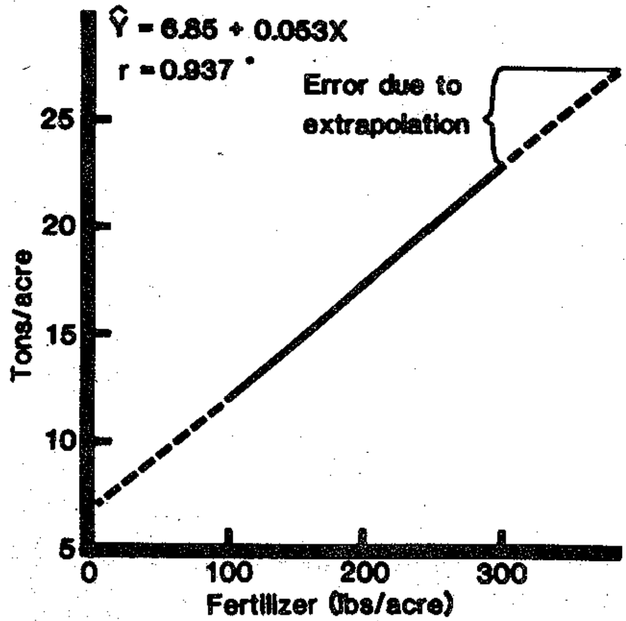

Fig. 3. Linear Regression

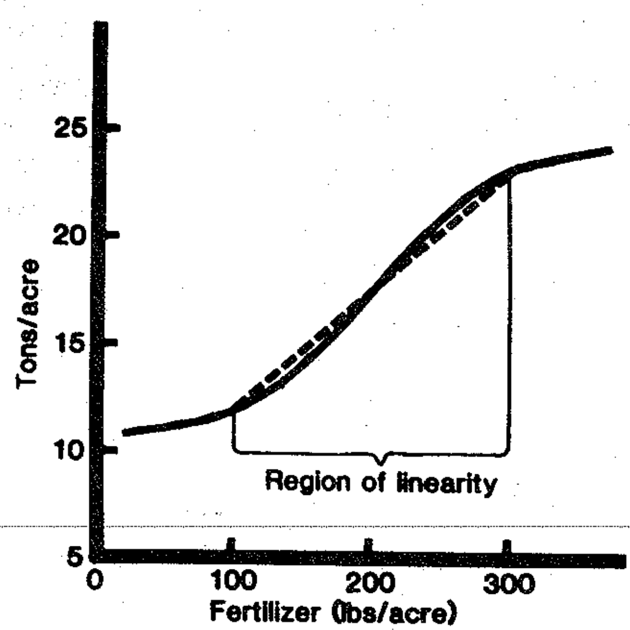

Fig. 4. Curvilinear Regression